Mr. George Shih
Resource Planning Dept.
South Florida Water Mgmt. Dist
P.O. Box V

155

George Shih

"TWO DIMENSIONAL INTERPOLATION AND UNCERTAINTY ESTIMATION BY AN INFORMATION-WEIGHTED-POINT-SPLINE METHOD"

DRE #116

Response Form

Author making presentation (Please note any changes):

Mr. George Shih
Resource Planning Dept.
South Florida Water Mgmt. Dist
P.O. Box V
West Palm Beach, FL 33402

To: Environmetrics 81 Program Committee
c/o SIAM/SIMS
1405 Architects Building
117 South 17th Street
Philadelphia, PA 19103

(Please place (✓) below as appropriate.)

( ) I have received your correspondence accepting my submission for presentation at Environmetrics 81 below:

155

George Shih

"TWO DIMENSIONAL INTERPOLATION AND UNCERTAINTY ESTIMATION BY AN INFORMATION-WEIGHTED-POINT-SPLINE METHOD"

(N.B.: An * appearing above denotes author making presentation)

(✓) I intend to make my presentation at Environmetrics 81 as planned.

( ) I prefer presentation format:

( ) Standard session

(✓) Poster session

(signed) George Shih

(date) Jan 28, 1981

TWO DIMENSIONAL INTERPOLATION AND UNCERTAINTY ESTIMATION BY AN INFORMATION-WEIGHTED-POINT-SPLINE METHOD*

George Shih

Given a set of irregularly located measurements, recently Kriging techniques have been shown capable not only to give estimated values but also to quantify estimation uncertainties. In this paper, the estimated value and the associated estimation uncertainty at an unmeasured point are attempted following a maximum likelihood approach by maximizing the information transferable from measured points to an unmeasured point. This is done by using a dynamic programming scheme. An estimated value is interpolated from a unique surface which is splined over the measured points according to their information weights. The estimation uncertainty is due to the variance propagation and the spline model deviations.

Maximum Likelihood and Information

From a single measurement at point i, it is reasoned that the best guess of the same variable at point o anywhere in the domain will be the same as the measured value, but the variance of the guess increases with distance. In case of multiple point measurements, a function passing near those measured points is considered to be a likely interpolation function of the variable. Hence the maximum likelihood approach is followed to specify the interpolation function.

For a given probability density function of f(Z), the likelihood function L of a state of nature Θ is given by (Lindgren, B.W. 1968, Statistical Theory)

$$(1)\quad L(\Theta) = f(Z_i;\Theta),\ i=1,\ldots N$$

where $Z_i(X_i,Y_i)$ are measured points

The information $I(\Theta)$ from a single observation Z_s is defined by

$$(2)\quad I(\Theta) = E\left[\frac{\partial}{\partial\Theta}\log f(Z_s;\Theta)\right]^2$$

Note that when observations are independent, information is additive.

Now let $\hat{Z}$ be estimated by an interpolation or spline function.

$$(3)\quad \hat{Z} = \Theta \cdot T$$

where Θ is a state vector and T is a position vector.

The maximum likelihood is achieved by setting the partial derivative of L with respect to each component of Θ to zero.

$$(4)\quad \frac{\partial L(\Theta)}{\partial\Theta_j} = 0 \qquad j = 1,\ldots k$$

where k is the number of components in Θ. There will be k equations to solve for k unknowns. When Z_i are independent normally distributed, then,

$$(5)\quad L(\Theta) = \left(\frac{1}{\sqrt{2\pi}}\right)^{-N}\left(\prod_{i=1}^{N}\frac{1}{\sigma_{io}}\right)\exp\left[\sum_{i=1}^{N}\frac{1}{2\sigma_{io}^2}(Z_i-\hat{Z}_i)^2\right]$$

$$(6)\quad I(\Theta) = \sum_{i=1}^{N}\frac{1}{2\sigma_{io}^2}$$

*For further information contact George Shih, South Florida Water Management District, P. O. Box V, West Palm Beach, Florida 33402

where σ_{io}^2 is the variance of using the Z_i value as an estimate at point o, and $\prod_{i=1}^{N}$ multiplication product of N terms.

Variance Propagation and Maximum Information

As mentioned before, it is reasoned that confidence in a guess decreases with the increase of distance. Suppose the propagation of variance is expressed by $\sigma_{io}^2 = \sigma^2(h_{io})$ where h_{io} is the distance between points i and o. Then σ_{io}^2 is the variance generated if the value measured at point i is used as an estimate of point o. When data values i and j are used to estimate for point o, the variances calculated from points i and j to point o, namely σ_{io}^2 and σ_{jo}^2 cannot be considered independent because of the intercorrelation between points i and j. In other words, to use two or more measurements to estimate an unmeasured point, the information derived from these measurements is not additive. The degree of correlation from j to i is

$$(7)\quad R_{ij} = 1 - \sigma_{ji}^2/\sigma_o^2$$

where σ_o^2 is the total variance at point o. Thus, to point o, information derivable from points i and j is

$$(8)\quad I^o = \frac{1}{2\sigma_{ic}^2} + \frac{1}{2\sigma_{jo}^2}(1 - R_{ji})$$

In order to obtain the maximum information from all the data points, the sequence of data point selection makes a difference. A dynamic programming approach is taken such that at each stage of single data point inclusion, the point with maximum additional information with respect to the existence of selected points is chosen.

$$(9)\quad I_n^o = \text{Max}\left[\frac{1}{2\sigma_{ko}^2}\prod_{j=1}^{n}(1 - R_{kj}\,]\right],\ k = n,\ldots N$$

where I_n^o is the maximum independent information contribution to point o due to inclusion of n^{th} point, selected among (N-n+1) available points with (n-1) points already selected. Because I_n^o diminishes quickly, the method results in a localized projection from a few nearby measurements for the estimation. When the dynamic programming principal holds the maximum information transferred from N data points to point o is

$$(10)\quad I^o = \text{Max}\sum_{i=j}^{N}\left[\frac{1}{2\sigma_{io}^2}\prod_{j=1}^{i-1}(1 - R_{ij})\right]$$

In other words, the minimum estimation variance achievable is $1/I^o$. It is easy to see when $\sigma_{io}^2 = 0$, $I^o = \infty$, point o can be computed without error. This occurs when point o falls exactly on one of the "error free" measurement points.

Estimation Uncertainty

The maximum information transferable from data points to the unmeasured point I^o is dependent on the variance propagation function selected, but independent of the spline model, or interpolation function to be selected. An imperfect spline model will further degrade the estimation reliability. In most practical

cases, little knowledge about the appropriate spline model is available. Statistical procedures, such as trend analysis, can be used to identify a significant function for interpolation. At any rate, a plane equation, which is the simplest model that can extend from point measurements to areal information, can be used. With respect to a given point o at (x_o, y_o), a unique set of σ_{io}^2, i=1,...N, determined according to the procedure of Equation (9), expands Equation (3) into three simultaneous equations to solve for $(a_o\ b_o\ c_o)$, the coefficients for the plane equation.

Note that each measured point i is weighted according to the information contributable to point o. If the selected spline model is perfect with this given set of data, it will pass through all the measurement points, e.g. $Z_i = \hat{Z}_i = a_o + b_o X_i + c_o Y_i$, i=1,...N. If the model is not perfect, the point discrepency is expressed by $(Z_i - \hat{Z}_i)^2$. Since each point is weighted, the effect of the model error to the estimated value is $\Sigma(Z_i - \hat{Z}_i)^2\ I_i^o/I^o$, where I_i^o and I^o are obtained from equations (9) and (10) respectively. Finally, the estimation uncertainty of Z_o is

$$(11)\quad U = \frac{1}{I^o}\left[\frac{1}{2} + \sum_{i=1}^{N}(Z_i - \hat{Z}_i)^2\ I_i^o\right]$$

Calibration of Variance Propagation Function

The variance propagation function is analogous to the variogram in Kriging. Consequently the function can be derived in the same way as the variogram. A variance propagation function is viewed more like a hypothetical function that is related to the data set, hence, it is easier to calibrate the function directly. The calibration procedure used here is sometimes called cross-validation.

Let the variance function be expressed by

$$(12)\quad \sigma^2(h_{io}) = \sigma_i^2 + f(h_{io}, \underline{P})$$

where σ_i^2 is the known measurement uncertainty of data point i, and f is a monotonically increasing function of distance h_{io} such as a power function, or a Gaussian function, with parameter vector $\underline{P} = (P_1, P_2, \ldots)$.

It is required that $f(0, \underline{P}) = 0$, so that when $\sigma_i^2 = 0$, measured point i possesses infinite information for that point. Otherwise the selection of f is very much judgmental. Note that $\sigma^2(h_{ij}) \neq \sigma^2(j_{ji})$ when $\sigma_i^2 \neq \sigma_j^2$

When a given set of measurement points are used in an interpolation, the implicit assumption is that this set of data is related to the unknown point. The maximum variance by propagation is,

$$(13)\quad \sigma^2 = f(h_{ij|max}, \underline{P})$$

where $h_{ij|max}$ is the maximum distance between the measurement points. This is considered to be the total variance σ_o^2 of all points to be used in Equation (7). At point o where $h_{io} > h_{ij}$, there may be $\sigma_{io}^2 > \sigma_o^2$, then the measurement at point i is considered to be useless for the estimation at point o. In other words, at a point away from the domain of data points, there may not be enough usable data to specify the interpolation function. An estimate at this point derived from this set of data is meaningless. Evidently, the selection of $h_{ij|max}$ as a range for information derivation is weak - any appropriate distance can be selected from judgment.

Cross-validation is a procedure to test a model performance by predicting the data themselves. Out of the N given data points, one of them, say i, is suppressed from the data list. Only the remaining N-1 points are used to estimate the value Z_i at i. For a given set of parameters $\underline{P}_k$ the sum of deviation squared $Q_k = \sum_{i=1}^{N}(Z_i - \hat{Z}_i)^2$ are recorded. The best set of parameters is the one that gives minimum Q_k, while matching estimation uncertainties on the entire region. That is, minimizes Q_k, subject to $Q_k = \sum_{i=1}^{N} U_i$, where U_i is the estimation uncertainty for data point i in the cross-validation.

Computation Procedures

1. Select a spline model, e.g. Equation (3).
2. Select a variance propagation function, e.g. Equation (12).
3. Calibrate for the parameters in Equation (12) by cross-validation. A simple algorithm was adopted for the purpose. To determine the θ in Equation (3), the weighting factors are determined according to Equation (9). The estimation value and uncertainty are computed by Equations (3) and (11), respectively.
4. With Equation (12) specified, Equations (3) and (11) can be solved at any (grid) point.

Discussion

To apply the method, three judgmental selections have to be made: the spline model, the variance propagation function, and the total variance. In each selection, the best choice is obviously dependent on the type of problem at hand. Nevertheless, the following observations may help in the selections.

It has been mentioned that a significant trend function may be used as a spline model. But what if the given data set is the residual of a trend function? It is known that each residual is prevailing only locally; on the entire region, the residuals sum up to be zero. Information for the support of the local variation is only derivable from a few nearby points. Hence a simple interpolation model, requiring few measurement points for identification, is preferred. In fact, it may be wise when a significant trend is identifiable, to work with the residuals using a simple plane as the interpolation model.

The variance propagation function is assumed to be monotonic, isotropic and homogeneous for simplicity in explanation. The fact that the method does not require $\sigma_{ij}^2 = \sigma_{ji}^2$ relaxes those limitations. The problems of using more complicated functions are associated with data support and physical process considerations. The behavior of the variance propagation function is more important in the small distance, up to around the mean distance between measurements. Fortunately, cross-validation procedures cover the range effectively.

The assumption of uniform total variance and its estimation are weak points in the present procedure. Further work is being done to incorporate the trend analysis into the method. The concept being that, after a significant trend is identified, the variance of the trend may be taken as the total variance at the point for the residual. Then the present procedure can be applied with given residuals and non-uniform total variances, to obtain the corrections on top of the trend function estimation.